



## Digitalizing Seismograms Using a Neighborhood Backtracking Method

Bi, Y., Feng, S., Zhao, G., & Han, B. (2016). Digitalizing Seismograms Using a Neighborhood Backtracking Method. In *Unknown Host Publication* (Vol. 9983, pp. 391-401). Springer.  
<http://uir.ulster.ac.uk/36203/1/KSEM2016.pdf>

[Link to publication record in Ulster University Research Portal](#)

**Published in:**  
Unknown Host Publication

**Publication Status:**  
Published (in print/issue): 05/10/2016

**Document Version**  
Author Accepted version

**General rights**  
Copyright for the publications made accessible via Ulster University's Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**  
The Research Portal is Ulster University's institutional repository that provides access to Ulster's research outputs. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [pure-support@ulster.ac.uk](mailto:pure-support@ulster.ac.uk).

# Digitalizing Seismograms Using A Neighborhood Backtracking Method

Yaxin Bi<sup>1</sup>, Shichen Feng<sup>1</sup>, Guoze Zhao<sup>2</sup> and Bing Han<sup>2</sup>

<sup>1</sup> School of Computing and Mathematics, University of Ulster  
Newtownabbey, Co. Antrim, BT37 0QB, UK

y.bi@ulster.ac.uk, Feng-S@email.ulster.ac.uk

<sup>2</sup> Institute of Geology, China Earthquake Administration  
Hua Yan Li, Chaoyang District, Beijing, China  
zhaogz@ies.ac.cn, zddhb@163.com

**Abstract.** In this paper, we present a new algorithm for tracing waves on seismograms and digitalising the waves into single vectors in the form of a time series data. The algorithm consists of two main components that will be used to handle the smooth and complicate cases, respectively. The underlying feature of the algorithm lies in a novel searching process based on examining the context of pixels to ascertain how the tracing moves forward. The algorithm has been evaluated on a limited number of samples and the result demonstrates its competence. The work presented can be regarded as an effort of developing a uniform earthquake archive covering both the historical and the digital device periods for future reassessment of the seismic hazard cross the world. The archive developed will serve as an effective means for discovering precursor of earthquakes by characterising the spectral seismograms and the source parameters of less active sources, whereby permitting comparative studies on earthquakes and development of prediction models.

**Keywords:** Analogous seismograms, digitalization of seismic wave, vectorization of analogous signal

## 1 Introduction

Many countries have a long history of using seismographs to record the ground motion with analogous measurements for monitoring earthquakes and place massive historical seismograms in the achieve storages [1]. Seismograms contain rich information including displacement, velocity, or acceleration of the ground motion, they are particularly important in zones with a low to moderate seismicity associated with seismogenic sources with medium to long recurrence intervals [2]. Although a large number of digital seismic networks have been constructed across countries in the past decades, there are no uniform seismogram archives that consist of both the historical and the digital device periods, allowing seismologists to conduct the reassessment of the historical seismic hazards. It is extremely difficult for scientists to retrieve original seismograms for a particular

historic seismic event and recover seismic information contained in old records [2]. Digitalisation of analogous seismograms is therefore imperative for developing uniform seismogram achieves, permitting researchers to analyse and study earthquakes at local and global scales with different historical periods.

Several studies have recently been reported in the literature [6], [7], [8], [9]. In [3], the authors developed an algorithm that can be used to extract waves from individual seismograms and concatenate them to 24 hour single waves. A further development of this work is a waveform tracing algorithm called Waveform Mosaic Algorithm that is able to approximately vectorize analogous seismograms based on local features in tracing seismograms [4]. An interactive system designed for automatic digitization of waveform has been reported in [5], the authors convert the digitization of seismograms as an inversion problem. The system integrates automatic digitization with manual digitization and users can interactively switch between these two modality functions to complete wave tracing tasks.

In this paper, we present a new algorithm for tracing waves on seismograms and digitalising the waves into single vectors in the form of a time series data. The development of this algorithm is based on the previous work [3], but the idea of this algorithm is on the basis of backtracking within a neighborhood of any dichotomous points (white and black), and concatenating black points in fragments into single waves in the form of digital vectors. The underlying difference between the algorithm from the previous one is, when searching next black points, all points within a neighborhood will be examined instead of the points in vertical lines based on local features. In addition, we develop an effective method for counting white points along the tracked waves, which can be used to evaluate and compare the performance of algorithms. This paper presents the detailed design of the algorithm and performance comparison with the previous algorithm [3].

Seismographs record the displacement, or velocity, or acceleration of the ground motion. Only one value is recorded at each time. No matter what the recording are, the traces aligned together and give several continuous series of black pixels, rather than one series, at each time point [2].

## 2 Description of Seismograph Readings

Seismometers, also called seismographs, are instruments that record motions of the ground, including those of seismic waves generated by earthquakes, volcanic eruptions, and other seismic sources. They could take measurements of ground motion and draw measurements on papers, each time only one value is recorded. No matter what the recording are, the traces are aligned together to form continuous series of black pixels. Due to the size of a paper, the continuous motion will be recorded on paper with equal segment of component waves for 24 hours, and then replaced with another paper. Fig. 1 illustrates an example of seismogram, the numbers labelled on the paper indicate recording times. In order to

vectorise seismograms, the first step process is to trace segments of waves and concatenate them into vectors in the form of a time series of format.

**Definition 1:** Let  $S_i = \{(x_0^i, y_0^i), \dots, (x_M^i, y_M^i)\}$  be a component wave on seismograms, where  $x_j^i$  is a time stamp and  $y_j^i$  is an amplitude value. In addition, let  $S = S_0 \tau S_1 \tau \dots S_K$  be a seismogram, where  $K$  is the number of component waves,  $\tau$  is an operation that concatenates component waves to a single seismogram.

From Fig. 1, we can see that it is straightforward to concatenate the time stamps of all component waves by operator  $\tau$  in a linear way, i.e.  $x_M^0 + x_0^1, x_M^1 + x_0^2, \dots, x_M^{k-1} + x_0^K$ , but the alignment of amplitudes to a single seismogram is complex. This will involve converting image pixels on a seismogram to amplitude values based on pixel grey-level values across waves, placing them onto the same coordinate system along with a zero point, and then adjusting them with offsets by accounting for the zero point. Simply, suppose  $y_0^i$  be a starting reference point of wave  $y^i$  in  $y$  coordinate axis and  $\Delta y^{i-1}$  be difference between  $y^i$  and  $y^{i-1}$  in  $y$  coordinate axes of two traced waves, then we have

$$\begin{aligned} y_j^{1'} &= y_j^1 - \Delta y_j^0, \\ y_j^{2'} &= y_j^2 - \Delta y_j^1 - \Delta y_j^0, \\ &\dots \\ y_j^{K'} &= y_j^K - \sum_{i=0}^{K-1} \Delta y_j^i \end{aligned}$$

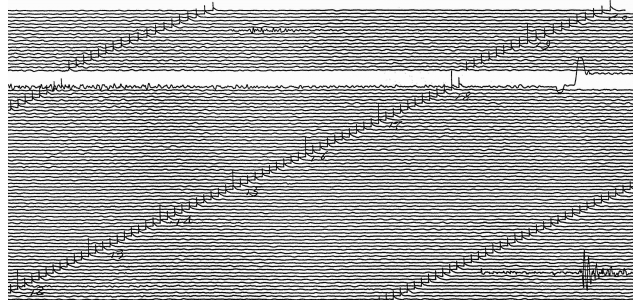
where  $0 \leq j \leq M$ . Following the above process, all component waves  $S_i (0 \leq i \leq K)$  will be placed down in the same  $y$  coordinate reference and then can be concatenated in a linear way corresponding to the concatenated time stamps. After that a seismogram will be vectorised into a time series wave. However prior to vectorising seismograms above, we have to perform image processing, i.e. a binarization process of enhancing image pixels with white and black points and use black pixels to trace each component waves into vectors. We detail the design of the algorithm in following sections.

As a seismogram wave can be either smooth or of fluctuation. The former does not contain much changes throughout the wave, whereas the latter may include many irregular parts with large variations. Thus the tracing algorithm consists of two main functions designed on the basis of smooth and complex situations.

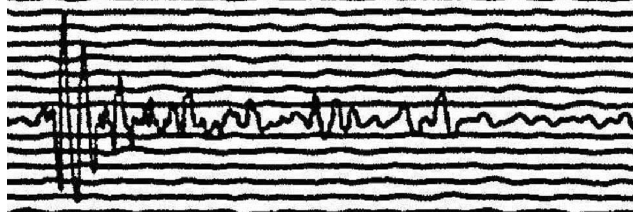
### 3 Description of Wave Tracing Algorithm – Tracing a Smooth Seismogram Wave

Zooming in a wave as shown in Fig. 2, we can take a fragment and illustrate it in Fig. 3. It can be seen that a wave is not a single pixel width line, actually it is composed of many pixels vertically. Instead tracing whole width pixel line,



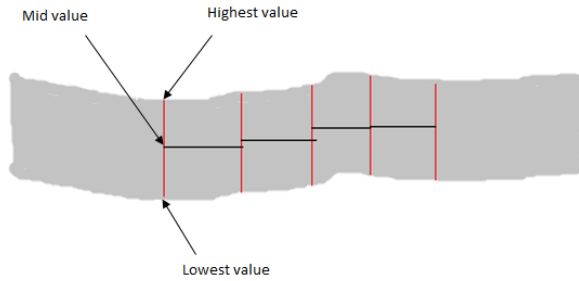


**Fig. 1.** A sample of original seismogram



**Fig. 2.** A fragment of the original seismogram in Fig. 1

the process is first to search the highest and lowest point values and then move forward in the middle of the wave pixel by pixel or with a number of pixels as a step.



**Fig. 3.** A single seismograph line

Specifically assume  $P_0(x_0, y_0)$  be the starting point of a wave, which is also referred to as the first key point,  $\varepsilon_x$  and  $\varepsilon_y$  be the searching steps. The searching process designed is, for a key point  $(x_i, y_i)$ , to move forward through checking whether the next point  $P_i(x_i + \varepsilon_x, y_i)$  is a black pixel. If yes, the process will vertically search from  $P_i(x_i + \varepsilon_x, y_i)$  up and down until reaching the highest

and lowest values, denoted by  $\mathcal{L}(y_i)$  and  $\mathcal{H}(y_i)$  as illustrated in Fig. 3, and then decide a new next key point  $P_{i+1}$  as follows:

$$x_{i+1} = x_i + \varepsilon_x = x_0 + (i + 1) \times \varepsilon_x \quad (1)$$

$$y_{i+1} = \frac{\mathcal{L}(y_i) + \mathcal{H}(y_i)}{2} \quad (2)$$

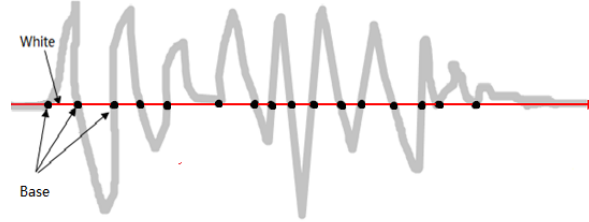
The above searching process is repeated until a new key point is not a black pixel and then turn to the complicate case given in the following section.

$$\omega = \mathcal{H}(y_i) - \mathcal{L}(y_i) \quad (3)$$

where  $\omega$  is the width of a wave, which will be used as a searching range for the complicate case.

#### 4 Description of Wave Tracing Algorithm – Tracing a Seismogram Wave with Fluctuations

Following the process given in Section 3, if  $P_i(x_i + \varepsilon_x, y_i)$  is not a black pixel, the tracing process becomes complex, and needs to account for the situations of irregular variations, wave crests and noise, etc.



**Fig. 4.** Searching key points in one pass

Given  $P_i(x_i + \varepsilon_x, y_i)$  being a white pixel, we assume  $V$  be a vector, holding so called base points. The first step of the searching process is to start from  $P_i(x_i, y_i)$  to search all base points as shown in Fig. 4 and put them into  $V$ . This process will result in  $V[0] = P_i(x_{i+0}, y_i), \dots, V[r] = P_i(x_{i+r}, y_i)$ .

For each base point, it divides a wave fragment to two parts – up and down ones, correspondingly the searching process consists of upward and downward cases. The following will detail how the searching process is conducted in tracing a wave into a vector.

#### 4.1 Upward Case

For each base point in  $V$ , by the smoothing algorithm the searching starts from a base point up and down vertically until reaching the leftmost and rightmost points that are referred to as the left and right slope points, as illustrated in Fig. 5. For each of these slope points, it is represented by a six by six grid, denoted by  $\mathcal{G}$ , as shown in Fig. 6, where the central cell  $\mathcal{G}_{[0][0]}$  holds either left slope or right slope point.

For the right slope point case of upward tracing, the searching process is composed of horizontal and vertical two steps. Through checking a white pixel in the upper right of the grid, the algorithm decides to turn the searching direction towards the leftmost point, denoted by  $\mathcal{L}_{\mathcal{M}}$ , and then the searching process will turn to move up. Prior to moving up, a key point has to be generated into a vector, denoted by  $V_{\mathcal{R}}^U$ . Specifically if  $\mathcal{G}_{[1][1]}$  is a white pixel, the search turns the left direction and apply the smooth algorithm horizontally trace until reaching the leftmost point as illustrated in Fig. 5, and then a key point will be generated into  $V_{\mathcal{R}}^U$  by Equation (4),

$$V_{\mathcal{R}}^U[i] = \left( \frac{|\mathcal{L}_{\mathcal{M}}.x - \mathcal{G}_{[0][0]}.x|}{2}, \mathcal{G}_{[0][0]}.y \right) \quad (4)$$

where  $i$  starts from 0 with one pixel increment. In addition, the searching from the right slope point to the leftmost point has to be stratified with an approximate certain range obtained by Equation (4), as it is very likely for one wave line to cross another, thereby searching the leftmost point might go astray.

Once the leftmost point reached and a key point generated, the second step is to use the leftmost point as a starting point for the smooth algorithm, the searching will move up until reaching the right slope point, which will be represented by  $\mathcal{G}_{[0][0]}$ . If  $\mathcal{G}_{[1][1]}$  is a white pixel, the searching towards the leftmost point will repeat the process as in the right slope case above until  $\mathcal{G}_{[-1][1]}$ ,  $\mathcal{G}_{[0][1]}$  and  $\mathcal{G}_{[1][1]}$  are all white pixels. This means the searching arrives at a crest of the wave, the searching process stop and the last key point will be obtained by Equation (5) below.

$$V_{\mathcal{R}}^U[i] = (\mathcal{G}_{[0][0]}.x, \mathcal{G}_{[0][0]}.y - \frac{\omega}{2}) \quad (5)$$

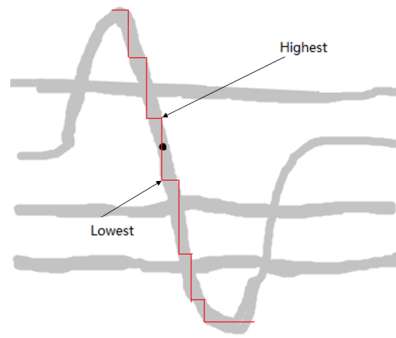
For the left slope point case, the searching process is almost the same as the right slope point case. Instead of checking  $\mathcal{G}_{[1][1]}$ , the algorithm will check if  $\mathcal{G}_{[-1][1]}$  contains a white pixel. If yes, the search turns the right direction and apply the smooth algorithm horizontally trace until reaching the rightmost point as illustrated in Fig. 5. Likewise a key point will be generated into  $V_{\mathcal{L}}^U$  by Equation (6),

$$V_{\mathcal{L}}^U[i] = \left( \frac{|\mathcal{R}_{\mathcal{M}}.x - \mathcal{G}_{[0][0]}.x|}{2}, \mathcal{G}_{[0][0]}.y \right) \quad (6)$$

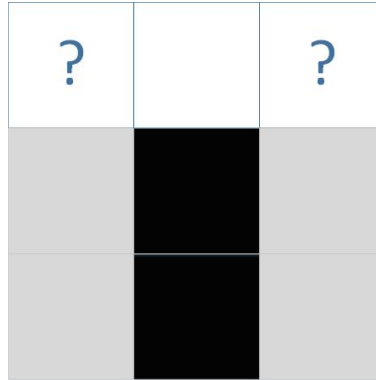
and the rightmost point will be treated as a starting point for the smooth algorithm, the searching will move up until reaching a left slope point, which will

be represented by  $\mathcal{G}_{[0][0]}$  again, the subsequent searching will repeat the above process until  $\mathcal{G}_{[-1][1]}$ ,  $\mathcal{G}_{[0][1]}$  and  $\mathcal{G}_{[1][1]}$  are all white pixels, that means that the searching arrives at a crest of the wave. In order to avoid generating a duplicate key point with the right slope case above, the searching process just stops and a key point will not be generated.

Once the searching process for one base point stop,  $V_{\mathcal{L}}^U[*]$  and  $V_{\mathcal{R}}^U[*]$  will be concatenated and then linked to  $P_i(x_i, y_i)$ . In the following, we describe the second part of the algorithm that the fragment wave is below the same base point.



**Fig. 5.** Tracing a curve in the vicinity of a key point



**Fig. 6.** Searching black pixels in the vicinity of a key point

## 4.2 Downward Case

In the downward case as illustrated in Fig.5, the searching process starts moving down vertically from the base point and move up when reaching a trough of the wave. The first point encountered is the left slope point, like the situation of the left slope point in the upward case above, this point is denoted by  $\mathcal{G}_{[0][0]}$ . If  $\mathcal{G}_{[1][-1]}$  is a white pixel, the searching will turn the right direction and trace horizontally until reaching the rightmost point, denoted by  $\mathcal{R}_{\mathcal{M}}$ , and then a key point will generated by Equation (7) below.

$$V_{\mathcal{L}}^D[i] = (\frac{|\mathcal{R}_{\mathcal{M}}.x - \mathcal{G}_{[0][0]}.x|}{2}, \mathcal{G}_{[0][0]}.y) \quad (7)$$

where  $i$  starts from 0 with one pixel increment.

The subsequent searching will turn from the rightmost point to move down vertically to find another the left slop point, denoted by  $\mathcal{G}_{[0][0]}$  again. This searching process will repeat until  $\mathcal{G}_{[-1][-1]}$ ,  $\mathcal{G}_{[0][-1]}$  and  $\mathcal{G}_{[1][-1]}$  are all white pixels, the searching processing stops and the last key point is obtained by Equation (5).

The second part of searching in the downward case is a moving upward process, which is similar to the situation of the left slop point in the upward case. In this situation, the left slop point is denoted by  $\mathcal{G}_{[0][0]}$ . If  $\mathcal{G}_{[0][0]}$  contains a white pixel, the searching will trace horizontally a rightmost point, denoted by  $\mathcal{R}_{\mathcal{M}}$ , and a key point will be generated by

$$V_{\mathcal{R}}^D[i] = (\frac{|\mathcal{R}_{\mathcal{M}}.x - \mathcal{G}_{[0][0]}.x|}{2}, \mathcal{G}_{[0][0]}.y) \quad (8)$$

where  $i$  starts from 0 with one pixel increment.

Noted that the downward case considers the left slop points in both moving up and moving down, which is different from those in the upward case. Here is a summary of the searching algorithm described above:

For the upward case:

- for the left slop point, turn right to reach the rightmost  $\mathcal{R}_{\mathcal{M}}.x$
- for the right slop point, turn left to reach the leftmost  $\mathcal{L}_{\mathcal{M}}.x$
- for the above two case, searching process is moving up

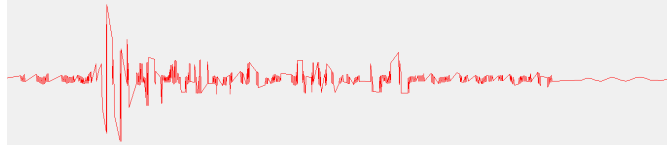
For the downward case:

- for the left slop point on the left hand side of a wave, turn to right and reach the rightmost  $\mathcal{R}_{\mathcal{M}}.x$ , and searching is moving down
- for the right slop point on the right hand side, turn to right and reach the rightmost  $\mathcal{R}_{\mathcal{M}}.x$ , searching is moving up

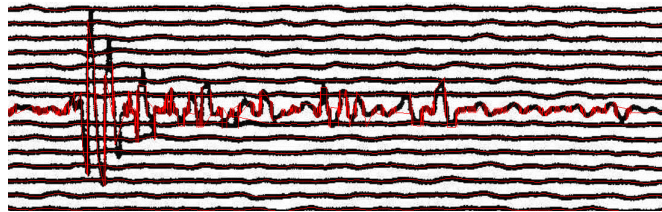
## 5 Evaluation - Comparing Traced Waves with Original Data

Given the algorithms in Sections 3 and 4, now we introduce how the accuracy of the algorithms for digitalising seismograms can be measured. Since there are no benchmark methods and data available for evaluation, we present the basic idea of the deigned measurement that can be used as an indicator of performance of the algorithm.

In the above, we use  $P(x_i, y_i)$  to hold all the traced points, where  $x_i$  represents time stamps on x-axis and  $y_i$  are wave amplitudes. The tracing algorithm picks  $y$  value with a ten pixel interval for the smooth case, and with less than ten pixels for the complex situation that can be tuned or adjusted. The assumption here is that each of  $P(x_i, y_i)$  corresponds to a black pixel on the original seismogram that was used to derive  $P(x_i, y_i)$ . However, if the corresponding pixel on the original seismogram is a white one, that means the tracing algorithm has not appropriately picked a black pixel. Fig.7 is a traced wave line, which has been redrawn on the original seismogram as shown in Fig.8. It can be seen that the traced wave approximately depicts the shape of the original wave seismogram, it is extremely challenging to trace the locus of seismic waves with precision and certainty. Fig.9 presents a tracing result by the previous algorithm [3].



**Fig. 7.** Tracing on a single wave line



**Fig. 8.** Tracing accuracy on single wave line

Additionally for each of the traced points in  $P(x_i, y_i)$ , the tracing process involves comparisons with all possible points within its neighbourhood. As illustrated in Fig. 6, the comparisons can involve  $2^3$  possible points to compare



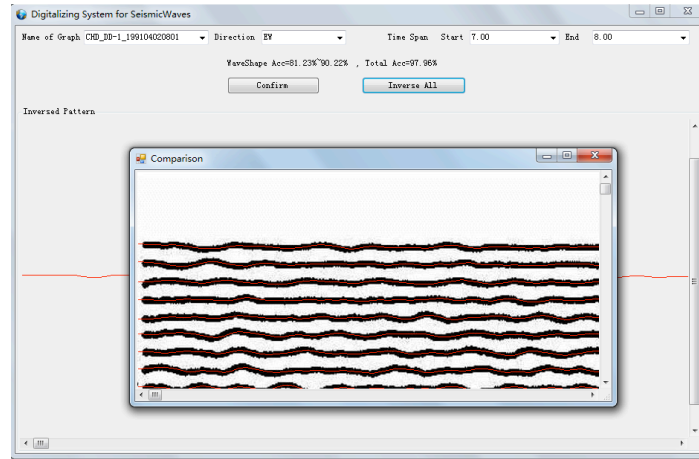
**Fig. 9.** Previous algorithm tracing on a single wave line

in order to ascertain tracing proceeding direction. Meanwhile from Definition 1, the number of windows is  $K$ , in which each window consists of  $M$  points. Therefore the computational complexity of the tracing algorithm can be estimated by  $O(K \times M + 2^3)$ .

In order to quantify the accuracy and make a quantitative comparison with the previous algorithm, we denote the number of white pixels held in  $P(x_i, y_i)$  by  $N_w$ , then the measure for tracing error is defined as Equation (9) below:

$$E_{\mathcal{T}} = \frac{N_w}{|P|} \quad (9)$$

where  $|P|$  is the length of vector  $P$ . By Equation (9), Fig. 10 illustrates an evaluation result, it can be seen that in a smooth case, 96% tracing accuracy has been achieved. Note that when calculating the accuracy, we did not consider the overlapping points, thus a range is added to estimate numbers of overlapping points on total number of points.



**Fig. 10.** An illustration of evaluation of the algorithm with 97% tracing accuracy

## 6 CONCLUSIONS

Although several attempts have been made and considerable improvements have been achieved in digitalizing analogous seismograms for the past decades, vectorised seismograms by automatic digitalization approaches are still not perfect, which could be reviewed as approximations of the original seismograms to some extent and require human intervention to subsequently undertake corrections. Even so, it is envisaged that the digitalization of historical seismograms allow research scientists to recover seismic information contained in old records, and perform comparative studies on the spectral characteristics and the source parameters of less active sources with the recent results. Therefore we believe that the algorithm developed in this study provides a useful tool for the digitalization of analogous seismograms and contribution to improvement of seismic hazard assessment in many regions of the world.

## 7 ACKNOWLEDGEMENTS

This work is partially supported by the project funded by the National Natural Science Foundation of China (Grant No. 41374077).

## References

1. Seismology Harvard Resources (<http://www.seismology.harvard.edu/resources.html#HRV>).
2. SanDron, D., Renner, G., Rebez, A., Slejko, D.: Early instrumental seismicity recorded in the eastern Alps. *Bollettino di Geofisica Teorica ed Applicata* 55 (4), 755-788 (2014)
3. WANG, M., JIANG, Q., FENG, J., FENG, S.: Researches on Key Algorithms in Analogue Seismogram Records Vectorization. *Sensors & Transducers*, 178 (9), 209-213 (2014)
4. Wang, M., Jiang, Q., Feng, J., Yu, X., Lin, N., Feng, S., Liu, N.: A New Waveform Mosaic Algorithm in the Vectorization of Paper Seismograms. *Sensors & Transducers*, 182 (11), 203-206(2014)
5. Xu, Y., Xu, T.: An interactive program on digitizing historical seismograms. *Computers & Geosciences* 63, 88-95 (2014)
6. Otsu, N.: A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man, and Cybernetics*, 9 (1), 62-66 (1979)
7. Xu, J., Xu, K., Wei, Y., Guo, Y.: Saving and data sharing of historical seismogram in Beijing National Earth Observatory, *Seismological and Geomagnetic Observation and Research*, 29 (3), 100-104 (2008)
8. Pan, Z., Feng, J., Wang, M.: A base-point searching algorithm in the digitization of seismograms, *Communications and Information Processing*, 289, 699-706 (2012)
9. Chen, B.: Research and implementation of log curve vectorization methods and rebuilding technique, Harbin Institute of Technology, (2007)
10. Cao, D., Tang, J., Wu, Y.: Algorithm for fast extracting human limb contours using searching NCM points, *J. Huazhong Univ. of Sci. &Tech. (Nature Science Edition)*, 35(5), 16-18 (2007)